

I/S: A JOURNAL OF LAW AND POLICY FOR THE INFORMATION SOCIETY

Foreword: Big Data Future and the First Decade of an Interdisciplinary Journal

PETER M. SHANE*

This last issue in *I/S*'s tenth volume and the forthcoming first issue in its eleventh represent a double milestone in the life of the Journal. They constitute, in published form, both much of the intellectual fruit of our most ambitious conference to date and a celebration of the Journal's first decade. The big data phenomenon provided an ideal focus for tenth anniversary festivities. It perfectly illustrates the twin convictions upon which *I/S* was founded. First, sometimes for better and sometimes seemingly for worse, the information and communications technologies of the digital age promise to reshape virtually every aspect of our collective economic, social, political, and cultural life. Second, understanding the appropriate roles of law and policy in meeting the challenges and opportunities that new technologies pose requires a wide-ranging interdisciplinary conversation. Lawyers and policy makers need to interact with scientists and engineers, social scientists, and humanists,

* Jacob E. Davis and Jacob E. Davis II Chair in Law, Moritz College of Law, Ohio State University. Big Data Future would not have happened without a grant from the Battelle Endowment for Technology and Human Affairs. I am indebted to my co-PI's on that grant for collaborating in the conceptualization and organization of the conference: Jennifer Evans-Cowley, David Landsbergen, Karen Lewis, Arnab Nandi, Beth-Anne Schuelke-Leech, and Caroline Wagner. We also received financial support from the Moritz College of Law Dean's Office and the College's Center for Interdisciplinary Law and Policy Studies, Ohio State's College of Engineering, and Elsevier. The Battelle Center for Science and Technology Policy in the John Glenn School of Public Affairs at Ohio State managed the conference web site and provided us the invaluable organizational services of its Administrative Associate – an understated title if ever there was one – Kelsey Chandler Nelson. Conference logistics were superbly managed by the *I/S* Symposium Editor for the 2013-14 Board, Julie G. Keys '14. Biographies of all speakers as well as videos of the panel discussions and keynote talks are archived at <http://bigdatafuture.org>.

both within and beyond the academy, if society is to shape its new tools to the maximum positive effect. The Big Data Future conference exemplified that interaction.

Marking *I/S*'s First Decade

The founding of *I/S: A Journal of Law and Policy for the Information Society* marked the happy convergence of a number of circumstances. For my part, I had just joined the Ohio State faculty after an intellectually challenging and deeply rewarding stint at what was then called the H. John Heinz III School of Public Policy and Management at Carnegie Mellon University.¹ Bringing technologists, social scientists, and humanities scholars together in interdisciplinary conversation about the social impacts of new information technologies had been my primary organizational assignment. That the world needed a strong journal at the intersection of law, policy, and information technology seemed obvious to me. Peter Swire, already a leading expert in privacy law and then on the Moritz College of Law faculty,² had been thinking similar thoughts about starting a privacy law journal, but was attracted to the idea of folding the privacy portfolio into the framework of a broader journal, which would publish one issue a year focusing on privacy and cybersecurity. The *Ohio State Journal of Criminal Law* had been founded just a year earlier, setting a happy institutional precedent for a project that combines traditional law review-style student editing with systematic peer review and faculty collaboration. Perhaps most important, Ohio State students interested, variously, in issues of intellectual property and cyber law had been lobbying over a period of years for the creation of a journal in the information law and policy space.

To be sure, the enterprise has had its frustrations – not the least of which has been our inability to come up with a “brand” more easily memorized than “*I/S: A Journal of Law and Policy for the Information Society*.” More seriously, the greatest difficulty of a self-consciously interdisciplinary journal is that university teachers and

¹ The Heinz School was distinctive among schools of public policy and management for the emphasis it placed on the management of IT-centered organizations. In 2007, the school's eminent information science faculty were organized as a distinct School of Information Systems & Management, which along with a reshaped School of Public Policy and Management, became part of the H. John Heinz III College at Carnegie Mellon University.

² In 2013, Professor Swire became the Nancy J. and Lawrence P. Huang Professor, in the Law and Ethics Program in the Georgia Institute of Technology's Scheller School of Business.

scholars do not generally secure tenure and other professional advancement as interdisciplinarians. They have to demonstrate the potential for intellectual leadership within disciplines, and that demonstration typically takes the form of publication in a limited number of prestigious within-discipline venues. Young law and technology scholars may even think it professionally advantageous to publish in the student-edited flagship journal of virtually any law school rather than in a peer-reviewed “specialty journal” targeting the audience they most hope to reach.

In addition, as a collaborative project with Carnegie Mellon University, *I/S* has realized only half its initial aspiration. Its very important success has been in luring a significant cohort of outstanding Carnegie Mellon faculty to serve as peer reviewers and, in some cases, as enthusiastic recruiters of interdisciplinary scholarship by graduate students, resulting in some of our most innovative offerings. Unfortunately, we have never figured out logistically how to engage Carnegie Mellon students as *I/S* staff members, which we had originally hoped would be a special attraction for our law student editors, as well.

Despite these bumps, our first decade’s authors³ have included communication scholars, computer scientists, economists, engineers, information scientists, journalists, political scientists, and sociologists – including such leading figures as Alessandro Acquisti, Pat Aufderheide, Danah Boyd, Andrew Chadwick, Stephen Coleman, Lorrie Cranor, Amitai Etzioni, Martin Libicki, Herb Lin, Jon Peha, and Vincent Price. Leading legal scholars published in *I/S* include Cary Coglianese, Rob Frieden, Michael Froomkin, Gillian Hadfield, Peter Jaszi, Jay Kesan, Ed Lee, Larry Lessig, Beth Noveck, John Palfrey, Pam Samuelson, Rebecca Tushnet, Jane Winn, and Christopher Yoo. Ohio State, alas, lost Peter Swire to the Georgia Institute of Technology, but Dennis Hirsch, the Geraldine W. Howell Professor at our neighboring Capital University Law School, has stepped in to help shepherd our annual efforts in privacy and cybersecurity.

Our annual conference confirms the breadth of our focus. We have held and published symposia⁴ on e-democracy, the mashup/remix phenomenon, youth and society media, cybersecurity

³ All *I/S* articles are available for free download in the “Archives” section of <http://www.is-journal.org>.

⁴ Links to materials from all our conferences are also available at <http://www.is-journal.org>, including videos of many of the lectures and panel discussions they featured.

policy, the future of online journalism, and competition and innovation in the broadband age. The current volume includes the first major academic symposium on the controversies surrounding foreign intelligence surveillance by the National Security Agency.⁵ Federal Communications Commission Chairman Tom Wheeler is set to keynote our spring, 2015 conference on “The Future of Internet Regulation.” And, of course, none of this would have been possible without the initiative and dedication of our student editors, many of whom have gone on to careers relating directly to the expertise they developed, in part, through *I/S*. This is the record, and the mission, that Big Data Future celebrates.

A Crash Course in the Big Data Phenomenon

A May, 2014, White House report tapped into much of the hope and uncertainty that surround the big data phenomenon. Noting that definitions of “big data” vary, the report cited a National Science Foundation solicitation that defines big datasets as datasets as “large, diverse, complex, longitudinal, and/or distributed datasets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future.”⁶ The amount of data that these sources yield staggers the imagination – authors both within this volume and elsewhere struggle to find the most jaw-dropping comparisons.⁷ But, as Farnam Jahanian notes: “The term ‘Big Data’ refers not only to the enormous volume of data being generated from a range of sources, but also its heterogeneity, complexity and diversity, as well as the rate at which it

⁵ The articles appear online at <http://moritzlaw.osu.edu/students/groups/is/volume-102/>.

⁶ Executive Office of the President, Big Data: Seizing Opportunities, Preserving Values 3 (May, 2014) [hereinafter EOP Report], citing National Science Foundation, Solicitation 12-499: Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA), 2012, <http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.pdf>.

⁷ The EOP Report, noting that 2013 estimates for the amount of global data generated reached 4 “zettabytes,” went on to explain: “A zettabyte is 1,000 000, 000, 000, 000, 000, 000 bytes, or units of information. Consider that a single byte equals one character of text. The 1,250 pages of Leo Tolstoy’s *War and Peace* would fit into a zettabyte 323 trillion times. Or imagine that every person in the United States took a digital photo every second of every day for over a month. All of those photos put together would equal about one zettabyte.” *Id.*

is generated.”⁸ These key aspects of big data are often summarized as “volume, variety, and velocity.”⁹

Because of our unprecedented capacity now to both gather data and to analyze it¹⁰, ambitions understandably loom large for potential improvements that data analysis can yield. The White House report seems modest in stating:

“[B]ig data analysis can boost economic productivity, drive improved consumer and government services, thwart terrorists, and save lives. . . . Big data technology . . . holds tremendous promise for better managing demand across electricity grids, improving energy efficiency, boosting agricultural productivity in the developing world, and projecting the spread of infectious diseases, among other applications.”¹¹

The title of an influential 2013 best-seller, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*,¹² probably comes closer to the spirit of most relevant media coverage. But – and this is a theme uniting many of the papers that follow – the big data phenomenon will live up to its revolutionary *positive* potential only under favorable technical, political, organizational, and intellectual conditions. These conditions cannot be taken for granted, but must be consciously pursued.

Issue 10:3 – Big Data Methods and Challenges

It was precisely for this reason that we asked Joel Gurin, a senior advisor to the Governance Lab at New York University and former editorial director and executive vice president of Consumers Union, to open the conference with a keynote address on “open data.” By “open

⁸ Farnam Jahanian, *The Policy Infrastructure for Big Data: From Data to Knowledge to Action*, 10 ISJLP 867 (2014).

⁹ EOP Report, *supra* note 6, at 4.

¹⁰ Though not without misgiving, I have adopted the convention of treating “data” as a singular noun and thus, as an “it.”

¹¹ *Id.* at 5-6.

¹² VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, *BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK* (2013).

data,” Gurin means data “intentionally released for public use,” in other words, “accessible public data that people, companies, and organizations can use to launch new ventures, analyze patterns and trends, make data-driven decisions, and solve complex problems.”¹³ Government agencies at all levels possess a lot of this high-value data, which means that (1) delimiting the scope of data that ought to be made public and (2) determining the appropriate investment in both money and effort to make that data public present significant issues of public policy. Mr. Gurin’s essay illustrates the current and potential benefits that come from open data and outlines a policy approach to open up more high-value data for public use. The challenge here, he argues, is not mainly technological: “Open Data is more of a philosophical movement, driven by the belief that data should be made available for public use on principle.”¹⁴

The first papers to follow the Gurin keynote are intended to highlight the variety of data under discussion and the difficulties posed in deriving actionable knowledge from data. From different angles, both within and beyond the academy, these authors highlight the challenges that big data poses for effective decision making at the individual or organizational level. Harvey J. Miller, a preeminent researcher at the intersection of geographic information science and transportation geography, documents how, through sensors, “[m]obility and connectivity are . . . generating massive amounts of data that can help manage” human activity.¹⁵ Deriving knowledge from such torrents of data poses massive technological challenges. But “[p]rocessing spatio-temporal data is only part of the story: we must also be able to describe it in a language that humans can understand. . . .”¹⁶ Professor Miller describes some of the methods through which this translation might happen, but notes: “The increasing speed at which data flows from the real world challenges our ability to understand and make decisions based on that data.”¹⁷

¹³ Joel Gurin, *Big Data and Open Data: How Open Will the Future Be?*, 10 ISJLP 861 (2014).

¹⁴ *Id.* at 704.

¹⁵ Harvey J. Miller, *Space-time Data Science for a Speedy World*, 10 ISJLP 705 (2014).

¹⁶ *Id.* at 711.

¹⁷ *Id.*

With her co-authors, Dr. Betsy Barry, a forensic linguist, addresses the phenomenon of text as data.¹⁸ The advent of electronic document storage has meant that lawyers engaged in civil litigation now encounter, in the process of discovery, an unprecedented amount of digitally searchable information. “When it comes to data,” they write, however, “quality eclipses quantity.”¹⁹ The authors call upon practitioners “to shift the focus of the conversation from size or quantity of data, to the quality of data, or from big data to better data, or even best data, as it were.”²⁰ Engaging in quality-driven discovery – focusing less on the volume of material disclosed and more on its relevance to the legal issues in dispute – will entail, according to the authors, the work of new interdisciplinary teams. That is because “[e]nsuring ESI quality in e-discovery requires a marriage of legal expertise, technical expertise, general subject-matter expertise and linguistic expertise.”²¹

For her part, Angela Shen-Hsieh, Chief of Product Design for IBM Business Analytics, counsels against mistaking the seeming precision of data as its chief virtue in generating actionable knowledge, especially through any automated algorithmic process.²² Not all data is equivalent in either trustworthiness or inherent interest. In the business world, she argues, data is most important in providing a context within which to understand and explore the narratives driving business decisions. Often, it is the relationships among different kinds of data – relationships which can helpfully be made comprehensible through visualization tools – that are more important than precise numbers. As Shen-Hsieh sees it, there is nothing in the world of big data that diminishes the need for acute human judgment. Her experience counsels that narrative, not data, ultimately drives decisions.

Dr. Angela Byers, who holds a Ph.D in Management Science and Engineering and is Director of Product Planning and Strategy in the Apps, Media, Publishing group at Microsoft, sees the big data

¹⁸ Betsy Barry et al., *The Big ESI: Going from Big to Better in E-discovery*, 10 ISJLP 719 (2014).

¹⁹ *Id.* at 720.

²⁰ *Id.*

²¹ *Id.* at 736.

²² Angela Shen-Hsieh, *In Defense of Imprecision: Humanizing Big Data for Business Decision Making*, 10 ISJLP 739 (2014).

phenomenon through a very different business perspective.²³ To be sure, “data, analytics, frontline tools, and people” must be brought “together to create business value.” The portrait Byers paints, however, is one in which the data seem to speak to decision making more precisely than in Shen-Hsieh’s account. She explains how businesses can use big data to:

- expose variability in performance and results;
- enable “experimentation, often involving rigorous statistics analyses to identify what option is better”;
- customize actions that target different subgroups of consumers;
- “supporting or even replacing human decision making with automated algorithms”; and
- enable new services that depend on highly granular and individualized information.²⁴

She cites “[e]merging academic research” suggesting “that companies that use data and business analytics to guide decision making are more productive and experience higher returns on equity than competitors that don’t.”²⁵

It may be that what the Shen-Hsieh and Byers papers reflect is analogous to a phenomenon that Eytan Adar, assistant professor in the School of Information & Computer Science and Engineering at the University of Michigan, spots in scientific research: a disjunction between a “social science community [that] has often focused on *explanatory models* (and *constructed data*)” and a research community in “the computational sciences [that] often target[s] *predictive models* (and *observational* or “*found*” data).”²⁶ He argues: “While Big Data veers towards the predictive/observational due to various structural reasons, it nonetheless offers significant benefits to

²³ Angela Byers, *Big Data, Big Economic Impact?*, 10 ISJLP 757 (2014).

²⁴ *Id.* at 759.

²⁵ *Id.* at 764.

²⁶ Eytan Adar, *The Two Cultures and Big Data Research*, 10 ISJLP 765 (2014).

those working in domains that demand theoretically grounded, explanatory models.”²⁷ He urges both camps to take the “opportunity to identify ways in which both modes of work can be used effectively in complementary ways.”²⁸

As Byers explicitly notes, however,²⁹ for society to realize the maximum benefits of our big data capacities, there are public policy issues that must be tackled that transcend the concerns of any one decision making organization. In his paper, Dr. Farnam Jahanian provides an excellent high-level summary of the big data phenomenon and its implications for “all facets of the discovery and innovation ecosystem, including the Nation’s academic, government, industrial, entrepreneurial, and investment communities.”³⁰ He cites a host of technical issues that the research and development community is already exploring with some intensity. But he focuses also on what has been the critical role of government in making “long-term, sustained investments in foundational computing, communications and information technology research, and [in] the development and deployment of large-scale facilities and cyberinfrastructure.”³¹ Government, he argues, has a part to play in workforce development “[i]n order to sustain technological advances of benefit to society.”³² Having served as head of the National Science Foundation Directorate for Computer and Information Science and Engineering, Dr. Jahanian, who holds the Edward S. Davidson Collegiate Professorship at the University of Michigan, offers a uniquely important voice in what is likely to be sustained debate over government’s roles. That debate is as much a political debate as a scientific one.

Rahul Telang, professor of information systems and management at Carnegie Mellon’s Heinz College, focuses his paper on perhaps the two most conspicuous policy challenges to the world of big data: privacy and cybersecurity.³³ In addressing these issues, noted

²⁷ *Id.* at 766.

²⁸ *Id.* at 767.

²⁹ Byers, *supra* note 23, at 757.

³⁰ Jahanian, *supra* note 8, at 870.

³¹ *Id.* at 872.

³² *Id.* at 883.

³³ Rahul Telang, *A Privacy and Security Policy Infrastructure for Big Data*, 10 ISJLP 883 (2014).

emphatically in the Byers³⁴ and Jahanian³⁵ essays as well, Professor Telang aims to identify the variety of options now confronting policy makers and their potential implications in trading off “positive benefits to . . . individuals, firms and society” against “the security and privacy threats” posed by the very data and analytic processes that promise those benefits.³⁶ He forecasts a greater need for public regulation where “firms are only indirectly responsible for consumer data,” as opposed to “markets where consumers deal directly with firms.”³⁷ It is perhaps unsurprising that Professor Telang believes that a key to sound policy making will be . . . more data.

Ashit Talukder, who leads and manages the Information Access Division in the Information Technology Laboratory at the National Institute of Standards and Technology (NIST), concludes this subsection with an essay on the importance of another kind of policy – namely, standards. “[S]tudies have shown,” he writes, “that available data is often not easily accessible and usable, which lowers the utilization of data and underlying knowledge.”³⁸ Addressing this problem, he argues, requires the development of standards that will foster the interoperation of big data systems, an objective understanding of the current state of big data science, and methods and tools for evaluating the utility of big data. The standards themselves may not represent the making of “public policy,” in the sense of social policy that trades off costs and benefits and attends to their distributional impacts. The public presumably does have a stake, however, in the development of sound standards through the collaboration of trusted institutions. This implies a robust role for a government agency such as NIST – now a sub-agency of the U.S. Department of Commerce – which is one of the nation's oldest physical science laboratories and a critical actor for over a century in meeting technology challenges central to U.S. economic competitiveness.

Returning to the theme of organizational structure and its relationship to the maximal success of data-driven decision making,

³⁴ Byers, *supra* note 23, at 763.

³⁵ Jahanian, *supra* note 8, at 881.

³⁶ Telang, *supra* note 33, at 786.

³⁷ *Id.* at 799.

³⁸ Ashit Talukder, *Big Data Open Standards and Benchmarking To Foster Innovation*, 10 ISJLP 801 (2014).

the last two papers in this first issue address “the roles, functions, and procedures within which an organization’s data is well managed and enabled as a strategic asset.”³⁹ Lawyer Barbara Cohn, New York State’s Chief Information Officer, calls the framework of protocols that determine those roles, functions, and procedures “data governance.”⁴⁰ Her essay offers a kind of agenda for assembling such frameworks, suggesting that they need to address issues within any data-driven organization of leadership, adaptability, structure, standards, and objectives.⁴¹ She provides several mini-case studies of organizations using data governance successfully to elicit actionable knowledge from large volumes of data.

Brian Cameron, the Executive Director of the Center for Enterprise Architecture in Pennsylvania State University’s College of Information Sciences and Technology, provides a yet more systemic overview of the data governance imperative through his introduction to the concept of “enterprise architecture.”⁴² The discipline of enterprise architecture, he explains, can do for organizations what the discipline of urban planning can do for cities.⁴³ He explains how its practice is evolving in many organizations from an exclusive focus on an organization’s information technology systems to a broader concern with the architecture of the entire organization to bring its goals and strategies into more effective alignment with its assets and processes.⁴⁴ The phenomenon of big data presents special challenges to contemporary enterprises, and he cites five common organizational failings that prevent enterprises from realizing maximum values from big data.⁴⁵ He concludes with a detailed analysis of how enterprise architecture can help avoid organizations to avoid these critical deficiencies.⁴⁶ That analysis amply demonstrates Professor Cameron’s

³⁹ Barbara Cohn, *Data Governance: A Quality Imperative in the Era of Big Data, Open Data, and Beyond*, 10 ISJLP 813 (2014).

⁴⁰ *Id.*

⁴¹ *Id.* at 815.

⁴² Brian H. Cameron, *The Need for Enterprise Architecture for Enterprise-Wide Big Data*, 10 ISJLP 831 (2014).

⁴³ *Id.* at 832.

⁴⁴ *Id.* at 833.

⁴⁵ *Id.* at 846.

⁴⁶ *Id.* at 852.

insistence that “Enterprise Architecture is a business issue, not a technology issue.”⁴⁷

Issue 11:1 – Big Data Impacts and Applications

The speakers on our conference’s second day by no means ignored the themes of challenge and complexity that dominated the first day. We organized the panels, however, to accentuate more concretely what big data could help society to accomplish. The papers written from these panels will appear in our next issue, which launches *I/S*’s eleventh volume.

As recounted by Scott Klein, the Senior Editor for News Applications at ProPublica,⁴⁸ one of data’s most venerable contributions to the public interest has been its deployment in what we would today call “accountability journalism.”⁴⁹ As a pioneer in the design of large interactive software projects that help to tell journalistic stories, Mr. Klein derives important lessons for data-driven journalists from Horace Greeley’s use of public data in 1848 to expose what appeared to be excessive reimbursements to members of Congress for travel between their home districts and Washington, D.C. The project definitely entailed what was “big data” for its day, but with predictable results in a pre-computing era. A table that Greeley’s co-author said involved over half a million figures was, as Klein explains, “rife with misspelled names, arithmetic errors, a missing entry and what must have been typographic errors introduced when typesetting the complex columns of numbers.”⁵⁰ Yet the story had impact: the House voted for reform.

Dr. Gary D. Bass, former director of OMB Watch (now the Center for Effective Government) and the current executive director of the Bauman Foundation, joins Klein in regarding open data as potentially key to enhancing government accountability, but argues also that government data needs to be managed within “a regulatory framework that protects the rights of the public and to serve as the

⁴⁷ *Id.* at 838.

⁴⁸ For an account of ProPublica’s origins and business model, see Richard J. Tofel, *Keynote Remarks: ProPublica and Journalism in the Public Interest*, 8 ISJLP 639 (2013).

⁴⁹ Scott Klein, *What a 166-Year-Old Data-Driven Story Can Teach Journalists Today*, 11 ISJLP 1 (forthcoming 2015).

⁵⁰ *Id.* at 6.

public's advocate in protecting our civil liberties and privacy.”⁵¹ He also stresses that the openness of information is often as or more critical than its volume:

“[T]ransparency for accountability includes Big and Little Data; it covers information about the government as well as the information it regulates and collects from companies and regulated entities. When it comes to transparency for accountability, size and complexity of the data is irrelevant: the key information used to promote accountability may be a single record or something said at a meeting.”⁵²

Decades of experience prompt Dr. Bass to anticipate ongoing resistance to such transparency from both within and beyond government. Experience and analysis also lead him to advocate urgent attention not only to data access, but also to the need for data standards to govern data quality and presentation. His essay concludes with an ambitious set of recommendations for strengthening accountability through data.

Another set of papers illuminates the potential for big data to serve the public interest with regard to health, education and welfare. The paper by Michael P. Johnson, Associate Professor in the Department of Public Policy and Public Affairs at University of Massachusetts Boston and an expert in operations research, is effectively a manifesto for a new sub-discipline in data analytics and information technology, namely, a combination of theory development and field research aimed at generating and evaluating “propositions related to capacity of CBOs [community-based organizations] to make productive use of data.”⁵³ He sets out a pilot field study of community-based organizations in the Boston area that illuminates the potential for his larger project. Drawing on this study, as well as a theoretical discussion, Professor Johnson concludes that “community-based organizations . . . have specialized needs for data analytics and IT, across multiple dimensions of organizational

⁵¹ Gary D. Bass, *Big Data and Government Accountability: An Agenda for the Future*, 11 ISJLP 13 (forthcoming 2015).

⁵² *Id.* at 20.

⁵³ Michael P. Johnson, *Data, Analytics and Community-Based Organizations: Transforming Data to Decisions for Community Development*, 11 ISJLP 49 (forthcoming 2015).

characteristics, that are not well addressed by applications designed for government, large or regional nonprofits, or for-profit organizations.”⁵⁴ Yet, he sees hope that further study will “enable students, researchers and practitioners to collaborate with community-based organizations to provide services, advocacy and knowledge rooted in data, analytics and information technology that can help . . . communities become desirable and sustainable places to live, work and visit.”⁵⁵

Nicholas Mader, a senior researcher at Chapin Hall at the University of Chicago, strikes a similar note in his essay, urging what he calls “In Situ research,” that is the use of administrative data regarding social programs that can usefully help to guide social policy by generating insights into the very populations, programs, and issues faced by policy makers in constructing such policy.⁵⁶ He argues that data drawn from all sorts of public agencies, “from school district offices to health and human services, to juvenile and adult criminal courts, to child and family welfare, to employment security, to city public health departments,” can be used to permit a kind of quasi-experimental analysis of social interventions that can significantly complement more traditional science methods for investigating social policy. Such studies, he believes, would offer significant new promise in evaluating the value of programs aimed at underserved populations, including low-income and minority youth.

Scott C. Burris, a professor of law and public health who directs the Center for Health Law, Policy and Practice at Temple University and the Robert Wood Johnson Foundation’s Public Health Law Research program, describes an ambitious project that perfectly captures the intersection of disciplines – the use of data to measure the intermediate and ultimate health impacts of laws and legal practices.⁵⁷ The enterprise required Professor Burris and his colleagues to figure out “transparent and consistent methods to transform the text of laws into numbers” so “the attributes of law – the independent variable – [would be captured] in a way that [would]

⁵⁴ *Id.* at 85.

⁵⁵ *Id.*

⁵⁶ Nicholas Mader, *The Big Data Era and an Integrated Mode of Inquiry for Social Policy-Relevant Research*, 11 ISJLP 97 (forthcoming 2015).

⁵⁷ Scott C. Burris, *Public Health Law Monitoring and Evaluation in a Big Data Future*, 11 ISJLP 115 (forthcoming 2015).

be accepted as reliable by the scientific community.”⁵⁸ Accomplishing this goal entails not so much high-tech analysis, according to Professor Burris, but rather laborious human coding of legal texts. The resulting data sets, however, may illustrate new tools for legal analysis that extend well beyond the public health context, creating “possibilities of policy surveillance and the integration of statutory and other policy data into the full range of platforms, informatics systems, mashups and uses that constitute our big data future.”⁵⁹

The final three papers in our second big data issue focus on big data opportunities and challenges for the future of cities. Professor Michael Batty chairs the Centre for Advanced Spatial Analysis at University College London and is a leading researcher in the mathematical modeling of cities. Like Harvey Miller, he urges that the paradigmatic “big data” in the context of cities comprises potentially unending data that streams from sensors that operate in real time.⁶⁰ Using a study dealing with such data from smart cards and geotemporal positioning in the public transport systems of greater London, he explores the difficulties of integrating such data with other data sets, which is necessary if the data are to be truly useful in helping to manage urban functions. Big data, for example, “often lacks the kind of structure that analysis requires so that the pattern and structure in such data can be exploited in terms of our understanding.”⁶¹ At least equally important, cities may well not have “the organizational structures . . . needed . . . to exploit big data and the analytics that is able to unpack them.”⁶² His study supports the view that “[t]he challenge with big data for the smart city is not simply technological, but more organizational and political.”⁶³

Along with his co-authors, Nader Afzalan, a geographic information systems lecturer and a Ph.D. candidate in urban planning and design at the University of Colorado, argues precisely in this vein that disaster management organizations and local governments

⁵⁸ *Id.* at 119.

⁵⁹ *Id.* at 125.

⁶⁰ Michael Batty, *Does Big Data Lead to Smarter Cities? Problems, Pitfalls and Opportunities*, 11 ISJLP 127 (forthcoming 2015).

⁶¹ *Id.* at 148.

⁶² *Id.*

⁶³ *Id.*

should develop an organizational capacity to gather and interpret big data generated through online social networks in order to facilitate on-the-ground recovery activities after natural disasters.⁶⁴ Their pilot study of fifty-two Facebook groups that were created after Hurricane Sandy reveals the potential for analysts, using a combination of content and network analysis, to identify much more effectively those local residents willing and able to take up key volunteer roles in the recovery process: “The organizations that are involved in disaster recovery should either gain the required skills for analyzing this complexity, or foster their collaboration with outside organizations to help them with this process.”⁶⁵

In her concluding essay, Sarah Williams, Assistant Professor of Urban Planning and the Director of the Civic Data Design Lab at MIT’s School of Architecture and Planning, points out that enthusiasm for the use of data to help determine city policies is not new, but that history shows frequent examples of data-driven models being misused in the policy process.⁶⁶ She argues that a combination of civically oriented strategies is available and necessary for working with data to create constructive action: “These include visualization, human-centered data collection strategies, sharing data, incorporating both quantitative and qualitative data analysis, acquiring private data to tell stories, and ultimately building diverse teams to tell narratives with data.”⁶⁷ She illustrates these strategies with stories of a number of exciting projects – the most remarkable of which for anyone who has visited Nairobi, Kenya is likely to be the highly collaborative initiative she describes to reduce traffic congestion and improve public service by generating a route map for “matatus.”⁶⁸ The uninitiated or naive observer of Nairobi’s streets might be forgiven for thinking that these privately owned minibuses operate with little regard for routes or timing. Several teams of researchers, however, were able to gather GIS data from drivers’ cell phones to create the

⁶⁴ Nader Afzalan, Jennifer Evans-Cowley & Maziar Mirzazad Barijough, *Social Networks and Natural Disaster Recovery: Integration of Online and On-the-Ground Activities*, 11 ISJLP 153 (forthcoming 2015).

⁶⁵ *Id.* at 171.

⁶⁶ Sarah Williams, *More Than Data: Working with Big Data for Civics*, 11 ISJLP 181 (forthcoming 2015).

⁶⁷ *Id.* at 185.

⁶⁸ *Id.* at 197.

first transit map of Nairobi, both verifying the existence of set routes and facilitating better scheduling.⁶⁹ Moreover, according to Professor Williams, the results of this and related efforts “have generated increased trust between our partners in Nairobi, new relationships with NGO’s and the Government, and engagement of Nairobi’s robust civic technology community, which has built applications on top of the data” her teams have gathered.⁷⁰ Her brief essay provides a rich introduction to strategies available to civic activists “for operationalizing data to expose issues and generate policy debates.”⁷¹

Concluding Reflections

Advances in data science and information technology have – as this conference well documented – laid the foundation for a future that could bring about a healthier, better educated, more productive, more sustainable, and better governed world than the one in which we now live. That science and that technology, however, have arguably laid the groundwork as well for a privacy nightmare⁷² and an unprecedented fragility of critical computer-dependent systems.⁷³

Big data enthusiasm also runs the risk of short-circuiting, rather than enabling thoughtful social analysis. As social researcher Kate McGraw has written:

“[H]ype becomes problematic when it leads to . . . “data fundamentalism,” the notion that correlation always indicates causation, and that massive data sets and predictive analytics always reflect objective truth. . . . Data and data sets are not objective; they are creations of human design. We give numbers their voice, draw inferences from them, and define their meaning through our interpretations. Hidden biases in both the

⁶⁹ *Id.* at 198.

⁷⁰ *Id.* at 190.

⁷¹ *Id.* at 181.

⁷² See generally *Symposium: NSA Surveillance: Issues of Security, Privacy and Civil Liberty*, 10 ISJLP 259 (2014).

⁷³ See generally *Symposium: Cybersecurity: Shared Risks, Shared Responsibilities*, 8 ISJLP 210 (2012), republished with additional material as PETER M. SHANE AND JEFFREY HUNKER, EDS., *CYBERSECURITY: SHARED RISKS, SHARED RESPONSIBILITIES* (2013).

collection and analysis stages present considerable risks, and are as important to the big-data equation as the numbers themselves. . . . We get a much richer sense of the world when we ask people the why and the how not just the 'how many.'"⁷⁴

In her essay, Dr. McGraw advocates greater interaction between data scientists and social scientists to deepen the understanding we get from data in the abstract. That recommendation echoes a theme permeating many of the papers in "Big Data Future" – the need for partnerships and collaboration across the boundaries of both organizations and fields of expertise to maximize the positive potential of big data and avoid the dystopian pitfalls. The need for multidisciplinary engagement with data in order to translate data into actionable knowledge echoes not only in our articles by academic researchers, but also in the essays contributed by our speakers from the world of business, journalism and government. Never before, it would seem, have the stakes been as high for society at large regarding our success or failure at generating such engagement.⁷⁵

Precisely for this reason, among the many positive reactions our conference elicited, my favorite was this one, from the director of an IT program management office:

"My expectations upon arrival were to gather information and perhaps hear an alternative viewpoint on big data as it relates to government. These expectations were quickly exceeded as one presenter after another illuminated various aspects of this timely subject. As an information technology manager, I regularly attend industry advisory briefings and I generally find these talks to be useful. The Big Data Future conference, however, provided a wonderful diversity of viewpoints that I didn't realize that I was missing. A journalist. A linguist. An attorney. A social

⁷⁴ Kate Crawford, *The Hidden Biases in Big Data*, HARVARD BUSINESS REVIEW BLOG NETWORK, (Apr. 1, 2013), <http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/>.

⁷⁵ In a forthcoming article, Professor Meg Leta Ambrose argues, based on the history of what she calls the "probalistic revolution" of the nineteenth century, that what will be transformative about big data is less likely to be the technologies and processes entailed in its acquisition and dissemination than the changes in though the availability of big data is likely to engender. Meg Leta Ambrose, *Lessons from the Avalanche of Numbers: Big Data in Historical Perspective*, 11 ISJLP (forthcoming 2015).

science researcher. An architect. *Each of these speakers focused a different light on issues I thought I understood.*⁷⁶

To my mind, the author's last line represents a complete validation for the intellectual vision that animated the creation of *I/S* ten years ago and the decision to celebrate that vision with an exploration of the big data phenomenon.

⁷⁶ Email to Prof. David Landsbergen from Brian D. Galloway, Interim Director of the IT Program Management Office, Ohio's Bureau of Workers' Compensation (Mar. 27, 2014) (on file with the author) (emphasis added).

